# Probing the self-association, intermolecular contacts, and folding propensity of amelogenin

**Moise Ndao,[1] Kaushik Dutta,[2] Keith M. Bromley,[3] Rajamani Lakshminarayanan,[3] Zhi Sun,[3] Gita Rewari,[1] Janet Moradian-Oldak,[3]* and John Spencer Evans[1]***

[1]Laboratory for Chemical Physics, New York University, New York, New York 10010

[2]New York Structural Biology Center, New York, New York 10027

[3]Center for Craniofacial Biology, School of Dentistry, Health Sciences Center, University of Southern California, Los Angeles, California 90033

Received 24 December 2010; Revised 2 February 2011; Accepted 4 February 2011
DOI: 10.1002/pro.603
Published online 23 February 2011 proteinscience.org

**Abstract: Amelogenins are an intrinsically disordered protein family that plays a major role in the development of tooth enamel, one of the most highly mineralized materials in nature. Monomeric porcine amelogenin possesses random coil and residual secondary structures, but it is not known which sequence regions would be conformationally attractive to potential enamel matrix targets such as other amelogenins (self-assembly), other matrix proteins, cell surfaces, or biominerals. To address this further, we investigated recombinant porcine amelogenin (rP172) using "solvent engineering" techniques to simultaneously promote native-like structure and induce amelogenin oligomerization in a manner that allows identification of intermolecular contacts between amelogenin molecules. We discovered that in the presence of 2,2,2-trifluoroethanol (TFE) significant folding transitions and stabilization occurred primarily within the N- and C-termini, while the polyproline Type II central domain was largely resistant to conformational transitions. Seven Pro residues (P2, P127, P130, P139, P154, P157, P162) exhibited conformational response to TFE, and this indicates these Pro residues act as folding enhancers in rP172. The remaining Pro residues resisted TFE perturbations and thus act as conformational stabilizers. We also noted that TFE induced rP172 self-association via the formation of intermolecular contacts involving P4–H6, V19–P33, and E40–T58 regions of the N-terminus. Collectively, these results confirm that the N- and C-termini of amelogenin are conformationally responsive and represent potential interactive sites for amelogenin–target interactions during enamel matrix mineralization. Conversely, the Pro, Gln central domain is resistant to folding and this may have important functional significance for amelogenin.**

*Abbreviations:* AQ-rP172, uniformly labeled $^{13}C$, $^{15}N$ rP172 in 90% v/v UDDW, 10% v/v $D_2O$, pH 3.8; CD, circular dichroism; DLS, dynamic light scattering; HSQC, heteronuclear single quantum coherence; IDP, intrinsically disordered protein; rP172, recombinant porcine amelogenin; 70-rP172, uniformly labeled $^{13}C$, $^{15}N$ rP172 in 30% v/v UDDW, 70% v/v 2,2,2-trifluoroethanol, pH 3.8; RC, random coil; SSP, secondary structure propensity score; TFE, 2,2,2-trifluoroethanol; UDDW, Milli-Q pure unbuffered deionized distilled water.

Additional Supporting Information may be found in the online version of this article.

Rajamani Lakshminarayanan's current address is Singapore Eye Research Institute, 7 Hospital Drive, Block C #02-02, Singapore 169611.

Grant sponsor: National Institute of Dental and Craniofacial Research-National Institutes of Health (NIH-NIDCR); Grant numbers: DE-020099, DE-013414; Grant sponsor: U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering; Grant number: DE-FG02-03ER46099; Grant sponsors: New York State Office of Science, Technology, and Academic Research, NIH; Grant number: P41 GM66354

**Correspondence to:* John Spencer Evans, Laboratory for Chemical Physics, New York University, 345 E. 24th Street, New York, NY 10010. E-mail: jse1@nyu.edu or Janet Moradian-Oldak, Center for Craniofacial Biology, CSA 107, School of Dentistry, Health Sciences Center, University of Southern California, Los Angeles, CA 90033. E-mail: joldak@usc.edu

**Keywords: biomineralization; amelogenin; enamel; folding propensity; nuclear magnetic resonance; intrinsically disordered proteins; calcium phosphate biominerals; dynamic light scattering**

## Introduction

Mammalian tooth enamel is one of the most highly mineralized materials of vertebrates.[1–3] The major protein component, amelogenin, is essential for normal enamel development and forms supramolecular assemblies (nanospheres)[4,5] that are believed to exert control over the mineral phase, morphology, organization, and directionality of hydroxyapatite crystal growth.[6,7] Recent bioinformatics and biophysical studies[8,9] have demonstrated that several amelogenins have sequence characteristics that fit the profile of intrinsically disordered proteins or IDPs.[9–14] The existence of unstructured regions within monomeric amelogenin have been confirmed by solution NMR studies of two recombinant amelogenins, murine[15] and porcine,[9] as well as model peptides representing amelogenin polyproline Type II (PPII) repeat regions.[16] In the case of the porcine species, under low pH aqueous conditions the monomeric form of amelogenin exists in an extended, unfolded state.[9] It is suspected that intrinsic disorder contributes to the structure and function of these proteins in self-assembly[5,9,15,17–22] and in cell-matrix, protein–matrix, and protein–mineral interactions.[5–9,23–25] The main challenge is to elucidate how intrinsic disorder affects the molecular behavior and configuration of this unusual series of proteins, and how amelogenin molecular behavior, in turn, affects biomineralization within the enamel matrix.

The current model of intrinsic disorder postulates that an unfolded protein possesses regions of disorder or residual structure that are recognized by targets for binding.[10–14] In some instances, once a given target binds to one of these regions, a conformational event may or may not accompany the binding event.[10–14] If we apply this principle to amelogenin, then the question arises as to which sequence regions would be conformationally attractive to potential enamel matrix targets such as other proteins or biominerals, or, to other amelogenin molecules during self-assembly. Recent NMR studies of recombinant porcine amelogenin (rP172) indicated that there are three primary regions that are structurally distinct.[9] These are the highly conserved N-terminus (P2–W45)[26] or the proposed self-assembly "A domain" (P2–M42),[20–22] the partially condensed, charged C-terminal domain (D155–D173),[9,20–22] and the extended Pro, Met, Gln-rich central domain (T58–P154)[9,16] that contains the polyproline Type II I70–P89 and P102–P145 sequence regions.[9] Both terminal domains exhibited evidence of residual secondary structure,[9] and these terminal regions may represent putative regions for folding during amelogenin–target interactions or self-assembly.[27]

It is experimentally challenging to assess the folding propensities of the three major regions of porcine amelogenin in the presence of matrix targets or during the self-assembly process. One of the major reasons for this is that the enamel matrix is a complex environment[2–6,20–25] and the true identity of amelogenin–specific targets, solution conditions, and other important features are not known at present. The other reason is that the amelogenin oligomerization process itself is quite complex[5,15,18,20–23,27] and results in the formation of high-order high-molecular weight complexes that are difficult to study with conventional structural techniques such as NMR. To circumvent this problem, we turn to techniques and approaches that can offer insights into the native folding of IDPs and create conditions that mimic the self-assembly process in a more tractable fashion. Over the last decade, the technique of "solvent engineering"[28] or the use of cosolvent mixtures of water with alcohols such as 2,2,2-trifluoroethanol (TFE) have been employed to change the environment around proteins and mimic certain conditions or to promote of protein folding and self-assembly with reasonable success.[28–33] Thus, we chose this approach to simultaneously achieve several goals that cannot be accomplished by other means with IDPs like amelogenin at the present time. These goals are: promote native-like structure within amelogenin, probe disorder-to-order transformations within the three major regions, and at TFE concentrations $> 40\%$ v/v[28–32] induce amelogenin oligomerization in a manner that allows identification of intermolecular contacts between amelogenin molecules.

For the first time, we report the use of TFE/water mixtures as an extrinsic "trigger"[28] to promote folding within rP172 amelogenin and simultaneously promote protein self-assembly[29] under conditions that normally favor the monomeric state (i.e., low pH, low ionic strength). These experiments reveal that both the N- and C-terminal domains undergo major folding transitions and form helical structures while the central domain is largely resistant to TFE folding. Concomitant to these events are changes in Pro backbone conformation at select sites within rP172, and the oligomerization of the protein via intermolecular contacts involving the hydrogen-bonding rich N-terminal region of the sequence. Thus, amelogenin is an IDP species that possesses alpha-helical folding propensities in the terminal regions of the protein molecule, and it is the

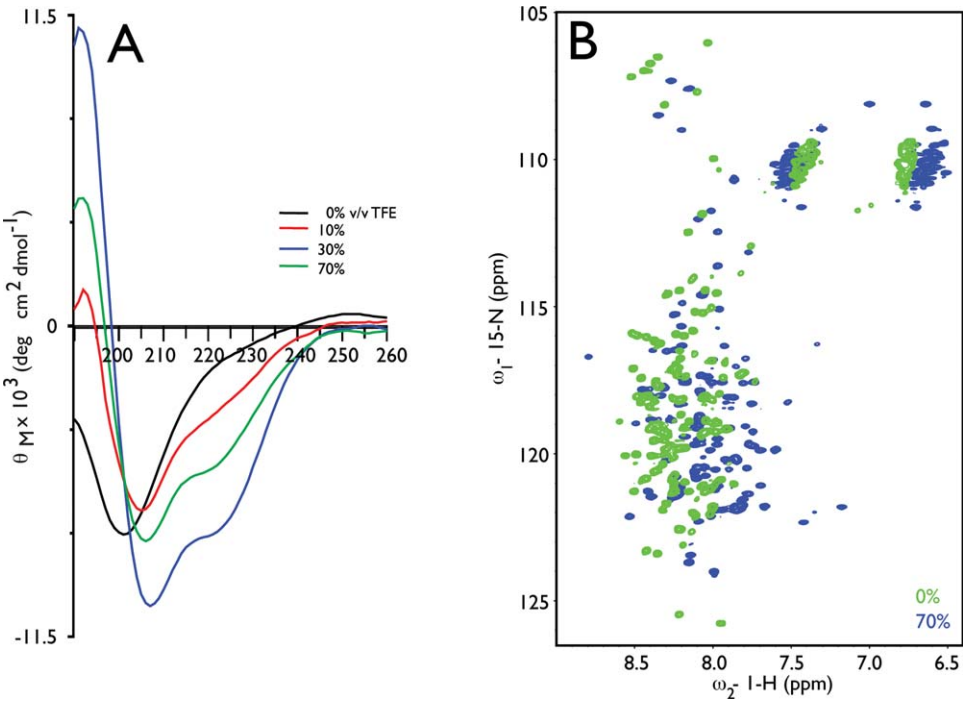


**Figure 1.** (A) CD spectra of unlabeled rP172, 0, 10, 30, and 70% v/v TFE in 50 m*M* Tris-HCl buffer, pH 5.5; (B) $^{15}$N–$^{1}$H HSQC overlay spectra of AQ-rP172 (green) and 70-rP172 (blue) at pH 3.8 in UDDW. HSQC spectra were obtained with 16 scans per increment, 1024 points in $\omega_2$ ($^{1}$H), 256 increments in $\omega_1$ ($^{15}$N), and a spectral window of 12 ppm in $\omega_2$ and 19 ppm in $\omega_1$. The size of the spectrum was 1024 by 256 points, with zero-filling in both dimensions. Sidechain and backbone fingerprint regions are shown.

N-terminal domain that plays a role in "solvent-engineered"[28] oligomerization.

## Results

### *The effect of TFE on the global conformation of rP172*

The amount of TFE required to induce maximum folding in a polypeptide is sequence-dependent.[28–33] For rP172, CD experiments performed at pH 5.8 in the presence of TFE reveal alcohol-dependent global conformational transitions [Fig. 1(A)]. The aqueous rP172 sample features only a single (-) pi–pi* transition band centered at 203 nm that is representative of a predominantly disordered protein that contains random coil (RC) and other secondary structure elements.[9] However, as TFE content increases, we note a corresponding increase in alpha-helical content as evidenced by the emergence of pi–pi* and n–pi* transition bands at 208 and 222 nm, respectively,[9] with a maximum effect achieved at 30% TFE. This CD dataset demonstrates that in the presence of TFE, certain regions within the rP172 protein fold into helical structures.

Evidence of TFE-induced global folding was also confirmed by NMR experiments [Fig. 1(B)]. For direct comparison with monomeric rP172, pH 3.8, in UDDW (AQ-rP172), we focused on the 70% v/v TFE rP172 sample, pH 3.8 (70-rP172) since this sample possessed significant TFE-induced folding effects [Fig. 1(A)] under conditions that sponsor oligomerization (i.e., TFE content $>$ 40% v/v).[28,29] A direct overlay comparison of the $^{15}$N–$^{1}$H HSQC AQ- and 70-rP172 spectra reveals resonant frequency mismatch arising from TFE solvent-exchange- and conformationally induced chemical shift effects.[28–35] Given that monomeric rP172 exists in a globally extended conformational state,[9] it is not surprising that this protein is susceptible to TFE -OH proton exchange[28–30] at NH backbone and sidechain sites. As a result, TFE and water proton chemical exchange significantly contribute to the differences in $^{1}$H frequency dispersion [Fig. 1(B)]. On the other hand, the $^{15}$N$\alpha$ spins, being directly bonded to both the carbonyl and alpha-carbon atoms of the peptide backbone, are sensitive not only to solvent-induced chemical exchange at NH sites but also conformational events brought about by the solvent-stabilization effects of TFE alcohol binding along the polypeptide backbone [Fig. 1(B)].[28–30] Hence, based upon the spectral difference between the AQ- and 70-rP172 samples, we conclude that TFE has a concentration-dependent effect not only on the solvent—rP172 NH chemical exchange rates but also the backbone conformation of amelogenin relative to the aqueous, monomeric state.

### *TFE induces rP172 self-assembly at low pH*

Observable backbone resonances were tabulated and assigned for the 70-rP172 sample (Figure S1, Table

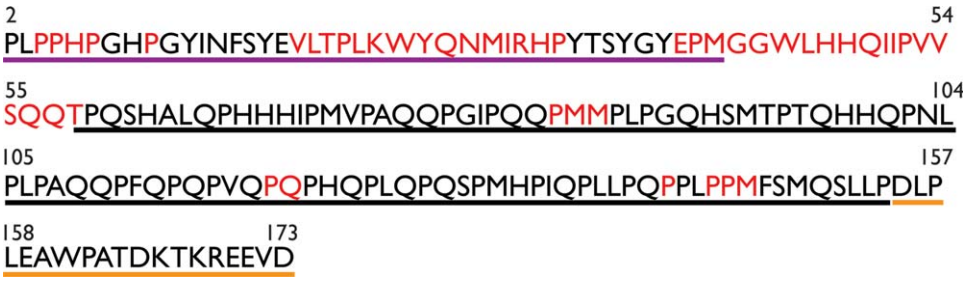


**Figure 2.** Primary amino acid sequence of recombinant porcine amelogenin (rP172). Residue M1 is missing from the recombinant version. Regions denoted in red represent undetected NMR resonances in the 70-rP172 $^{15}N$–$^{1}H$ HSQC spectra. Underlined regions denote the highly conserved N-terminal or self-assembly A domain (purple), the PPII-containing central domain (black), and the charged hydrophilic C-terminal domain (orange).

S1, Electronic Supporting Information, note assignment protocol in table legend). A surprising finding was the effect of TFE on the number of observable HSQC crosspeak resonances. Relative to the AQ-rP172 sample where only the E40, Q126, and H132 resonances are undetected,[9] the 70-rP172 sample possess significantly fewer assignable HSQC crosspeaks and we were able to assign all but 45 residues (representing 26% of the total sequence) (Fig. 2; Figure S1, Table S1, Electronic Supporting Information). Interestingly, the majority of these missing residues (82%) lie within N-terminal sequence blocks P4–H6, V19–P33, and E40–T58 (Fig. 2). From previous NMR studies of salt, Ca (II)-induced amelogenin dimerization, it is known that the reduction in intensity and attenuation of $^{15}N$–$^{1}H$ HSQC crosspeaks occur due to conformational exchange.[15] Selective loss or broadening of NMR signals was also observed in aggregation studies involving various proteins in buffer salt environments[34] or TFE.[28] Thus, the attenuation of HSQC crosspeaks reflects TFE-induced protein–protein interactions involving intermolecular contacts with residues in the P4–H6, V19–P33, and E40–T58 regions of the N-terminus, and to a minor extent, short sequence regions (P84–M86, P141, P144–P145) within the central domain (Fig. 2, Figure S1, Electronic Supporting Information). These interactions lead to changes in backbone dynamics, intermediate time scale broadening, and signal intensity attenuation of the HSQC resonances that are involved in intermolecular contacts.[34] Coincidentally, the N-terminal regions T21–R31 and Y12–I51 of mouse amelogenin underwent similar intermediate time scale motion broadening during counterion-induced dimerization,[15] and this correlates with our present findings.

Other features of the $^{15}N$–$^{1}H$ HSQC spectra provide additional evidence that TFE affects rP172 protein backbone dynamics. Specifically, we note the presence of unidentified $^{15}N$–$^{1}H$ HSQC resonances in 70-rP172 that are associated with conformational exchange processes (Figure S1, Electronic Supporting Information). In AQ-rP172, 4 conformational exchange peaks (V54, A63, W161,T164) were observed,[9] but in the 70-rP172 sample we identified 15 conformational exchange-related crosspeaks, a three-fold increase. We believe that this alteration in protein conformational exchange occurs in response to a number of factors, including TFE solvent–peptide backbone interactions and dynamics[28–33] and rP172 self-association.

To verify TFE-induced self-assembly, we analyzed the distribution of amelogenin particle sizes at pH 5.5 in 0% and 10% v/v TFE (1 mg/mL rP172; Fig. 3). Note that it was necessary to perform the experiments at pH 5.5 and at lower TFE concentrations to be able to reliably measure the hydrodynamic radius of amelogenin monomers.[17,18,27] If the pH is too acidic, the positive charge on the surface of the monomers can cause accelerated diffusion due to charge-charge repulsion.[17,18,27] Similarly, we found that, despite thorough mixing, TFE content $>$ 10% v/v increases light scattering and leads to errors in interpretation. The introduction of 10% v/v TFE lead

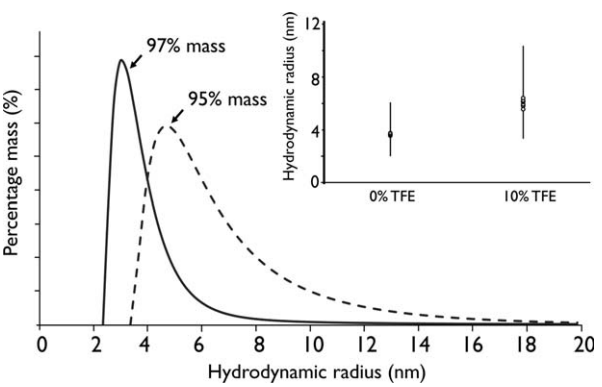


**Figure 3.** A comparison of the size distributions of rP172 (1 mg/mL, pH 5.5, 22°C) in the absence of TFE (solid line) and in the presence of 10% v/v TFE (dashed line). Inset figure is a plot of $R_H$ versus % v/v TFE showing the increase in $R_H$ from 3.66 to 6.04 nm. The error bars represent the skewed polydispersity of the particles. Multiple measurements were performed for each sample, revealing low standard deviation values for both the 0% and 10% v/v TFE samples (0.06 and 0.26 nm, respectively).

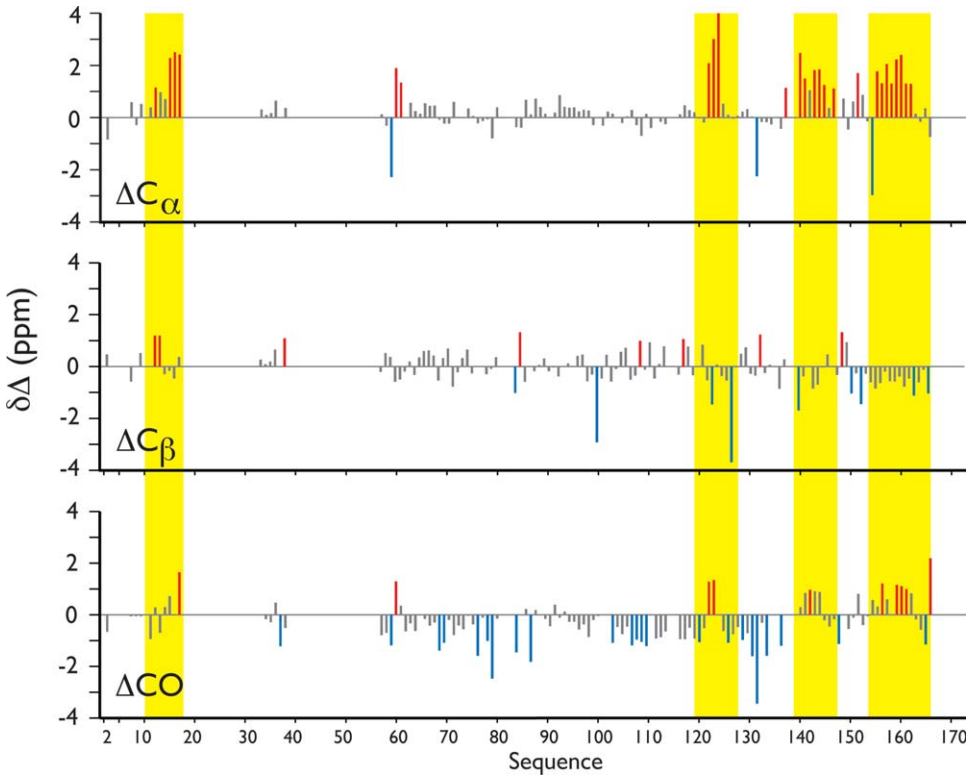


**Figure 4.** $^{13}C\alpha$, $^{13}CO$, $^{13}C\beta$ random coil conformational shifts for 70-rP172. Values were determined using the NMR data presented in Table S1 (Supplementary Information) and $^{13}C$ NMR random coil chemical shifts obtained from protein database values.[36,37] Yellow stripes denote sequence regions containing three or more contiguous residues exhibiting nonrandom coil characteristics. Regions without corresponding histogram bars represent TFE-induced relaxationally broadened resonances (Fig. 2; Figure S1, Electronic Supporting Information). In general, (+) secondary shifts for $^{13}C\alpha$ and $^{13}CO$ indicate helical structures whereas (−) shifts indicate a propensity toward extended conformations such as beta-strand.[36,37] Red histogram bars denote (+) $^{13}C\alpha$ and $^{13}CO$ conformational shifts $>$ 1 ppm (alpha-helix) for residues Y12–Y17, H62–A63, Q91, Q109, L125, P127–S129, M146–M149, S151, L153–D155, L158-E159, P162–R169. Blue histogram bars denote (-) $^{13}C\alpha$ and $^{13}CO$ conformational shifts $>$ 1 ppm (beta-strand) for residues P2, S61, Q79, P87, Q101, P105, Q110, Q113, P114, Q123, L137–P139, P143, Q150, W161. For $^{13}C\beta$, (−) values indicate helical conformations and (+) values indicating extended conformations.[36,37]

to an increase in the average hydrodynamic radius of rP172 (i.e., from 3.66 ± 1.69 nm in the absence of TFE to 6.04 ± 2.25 nm in the presence of 10% TFE). While the percentage of polydispersity in both samples is the same (around 45%), the presence of larger oligomers with an $R_H > 8$ nm increased significantly in the 10% TFE sample. Hence, our NMR and dynamic light scattering (DLS) data confirm that TFE induces self-assembly of amelogenin, even at the relatively low concentration of 10% v/v.

### *Localization of folding within rP172*

Since αNH and $^{15}N\alpha$ chemical shifts encode solvent-related effects,[10,11,34,35] we analyzed sequence-specific backbone $^{13}C$ chemical shift information (i.e., $^{13}C\alpha$, $^{13}C\beta$, $^{13}CO$) to identify regions which experience TFE-induced folding. We performed two analyses. The first involved the calculation of 70-rP172 $^{13}C$ backbone conformational shifts relative to protein database values obtained for the random coil state (Fig. 4).[36,37] The second involved a calculation of 70-rP172 $^{13}C$ backbone conformational shifts relative to the monomeric AQ-rP172 low pH aqueous sample from our previous NMR study (Figure S2, Electronic Supporting Information).[9] In general, (+) secondary shifts for $^{13}C\alpha$ and $^{13}CO$ indicate helical structures, whereas (−) shifts indicate a propensity toward extended conformations such as beta-strand.[9,36,37] For $^{13}C\beta$, (−) and (+) values indicate helical and extended beta strand conformations, respectively.[9,36,37]

If we define significant deviation in backbone conformational shifts to be $\geq$ 1 ppm[9,36,37] then we can identify where significant TFE-induced folding transitions occur within the rP172 sequence (Fig. 4, note yellow stripes). We observe four sequence clusters where consistent (+) conformational shift deviations (alpha helix) involve three or more contiguous residues: (a) Y12–Y17; (b) L125–S129; (c) M146–D155; and (d) L158–R169. These deviations were mirrored in the data comparisons for 70-rP172 versus AQ-rP172 (Figure S2, Electronic Supporting

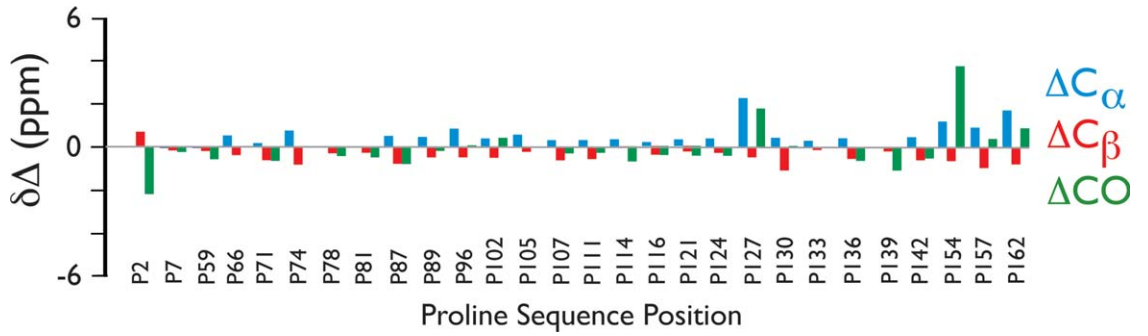


**Figure 5.** $^{13}$C backbone and $^{13}$Cβ sidechain conformational shifts obtained for observable Pro residues within 70-rP172 relative to published values obtained for these same residues within AQ-rP172.[9] Data was obtained from analyses performed in Figure S2, Electronic Supporting Information.

Information). From our previous studies with the low pH monomeric form of rP172, we know that regions Y12–Y17, M146–D155, and L158–R169 adopted a random coil conformation, and, region L125–S129 adopted a PPII conformation.[9] Hence, TFE induces alpha helical folding transitions within these regions. We also identified limited sequence regions within the PPII central domain where (−) CO conformational shifts were noted (Fig. 4), indicating a folding transition to an extended beta strand state in these regions. Similar phenomena were observed in other PPII sequences, i.e., TFE-induced loss of PPII structure.[38] However, given that only limited regions within these two PPII domains experience TFE-induced folding transitions, it is clear that the majority of these Pro, Gln rich regions within the central domain are resistant to TFE "solvent engineering," most likely as a result of the strong stabilization features of these sequences.[39,40]

### *Conformational response of Pro residues*

It is known from earlier studies that TFE affects Pro imido ring dynamics, conformation, and *cis trans* interconversion.[28,29,33,38] Since amelogenins contain a significant number of Pro residues (e.g., rP172 = 40 Pro, or 23% of the total sequence, Fig. 2), we would expect a globally extended, solvent-accessible molecule like rP172 to exhibit significant TFE perturbation of global Pro conformation and dynamics. Surprisingly, this is not the case. As shown in Figure 5, we analyzed the $^{13}$Cα, $^{13}$Cβ, and $^{13}$CO chemical shifts obtained for the 70-rP172 versus those published for AQ-rP172 (Fig. 5).[9] Using the criteria of significant conformational shift deviation to be $\geq 1$ ppm[9,36,37] we find that only a limited set of seven Pro residues (P2, N-terminus; P127, P130, P139, P154, central domain; P157, P162, C-terminus) respond to TFE, with the largest effects observed within the C-terminal half of this protein. Since the Pro imido ring Cα and Cβ chemical shifts are sensitive to polypeptide backbone conformation,[38,41,42] we infer that TFE has a select effect on the conformation of these particular prolines, but that the remaining detectable Pro residues within rP172 are resistant to TFE-induced conformational change.

### *Comparison of aqueous monomeric and 70% v/v TFE rP172 secondary structures*

We compared the secondary structure probabilities for observable 70-rP172 sequence regions alongside the full sequence regions of AQ-rP172,[9] using the secondary structure propensity (SSP) scoring system for IDP proteins[43] and by analyzing the $^{13}$Cα and $^{13}$Cβ chemical shifts (Table S1, Electronic Supporting Information). A SSP score of +1 or −1 at a given residue reflects fully formed alpha-helix or beta-strand, respectively, while a score of 0.5 indicates that 50% of the conformers in the disordered state ensemble are helical at that position.[43] As shown in Figure 6, AQ-rP172 does not feature any fully formed helical or beta strand as evidenced by SSP scores $< 1$. However, if we examine the (+) or (−) SSP trend, we observe that a greater probability of

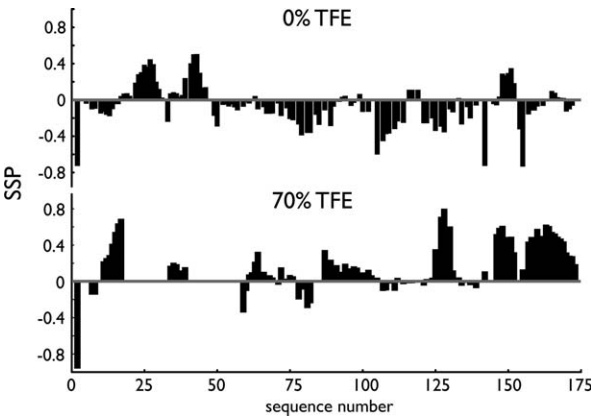


**Figure 6.** Secondary structure propensity scoring (SSP) for AQ-rP172 (0% TFE) and 70-rP172 (70% TFE). Data for the monomeric AQ-rP172 state was taken from ref. 9. In the 70% v/v sample, sequence positions where histogram bars are absent denote residues that experience intermediate time scale broadening and are absent from multidimensional NMR spectra.

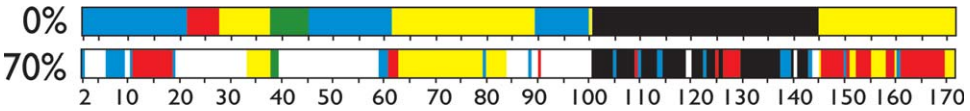


**Figure 7.** Comparison of secondary structures obtained for AQ-rP172 and 70-rP172. Data for the monomeric state was taken from Ref. 9. In the case of the 70-rP172 sample, secondary structures were estimated using $^{13}C\alpha$, $^{13}C\beta$, and $^{13}CO$ conformational shifts (Fig. 4). Red = alpha helix; Blue = extended beta strand; white = missing residue(s); Black = PPII; Yellow = RC.

alpha helix exists within the N-terminus (I13–Y26, N28–Y39) and limited regions of the central domain (Q113–Q115; M146–S148) and the C-terminus (T164–K168). Conversely, a higher probability of beta strand is observed in P2–Y12 of the N-terminus, Q49–F112; V117–P142; M149–P157 of the central domain, and L158–P162; R169–D173 of the C-terminus.

When we compare these structural probabilities against those obtained for 70-rP172, we observe that the addition of TFE leads to overall higher SSP scoring (0.5 to 1.0), indicating that more stable secondary structures are forming. Moreover, it is clear that TFE induces a beta structure-to-alpha helical conversion throughout the rP172 sequence, with the highest alpha helical probabilities observed for P2–W45 and P145–V172. Hence, our SSP calculations indicate that rP172 molecule possesses residual beta strand structures under monomeric conditions, and it is the terminal regions which transform into a predominantly helical structure in the presence of TFE.

These findings also correlate with secondary structure calculations using the $^{13}C\alpha$ and $^{13}C\beta$ conformational shifts obtained for AQ-rP172[9] and 70-rP172 (Fig. 7). Where the deviations were minor compared with the published conformational shifts for the low pH, monomeric form (Figure S2, Electronic Supporting Information), we elected to retain the original secondary structure classifications.[9] Obviously, since a number of residues cannot be identified due to putative intermolecular associations (Fig. 2), we consider this approach to be realistic given these limitations. The main difference between the two rP172 samples is the emergence of alpha-helical conformation in 70-rP172, which supports the results obtained from CD TFE studies [Fig. 1(A)] and the SSP calculations (Fig. 6). Specifically, the principle locations of TFE-induced helical structure are within the extended beta strand regions of the N-terminal region (Y12-Y17) and within the random coil C-terminal region (P145-D173). It appears that these are the sequence regions where TFE exerts its strongest effects. Second, we note that, with the exception of short helical and extended beta strand structure transitions, a significant portion of the PPII-containing central domain (I70–P89; P102–P145) is unaffected by TFE content.

## Discussion

Amelogenin protein plays a critical role in controlling tooth mineralization and extracellular matrix signaling activities.[24,25] There are a plethora of different targets that engage amelogenin during biomineralization, such as other amelogenin molecules[4,5,17,18,20] biominerals[6,7,23] cell receptors,[24,25] and other enamel matrix proteins such as the 32 kDa enamelin.[22,44] The ability to interact with such a diverse array of targets may be related to the "conformational plasticity" of amelogenin, i.e., the tendency of specific regions to adopt a specific configuration when needed.[8,9,12–14] The present "solvent engineering"[28] study reveals that the primary candidates of "conformational plasticity" within amelogenin are the N- and C-termini (Figs. 4–7). Both regions form unstable residual structures in water,[9] but when their environments are altered by "solvent engineering," their structural stabilities increase (Fig. 6), helical structures form (Figs. 4, 6, 7), and detectable global folding occurs within the rP172 amelogenin molecule (Fig. 1). Based upon these observations, we propose that these terminal domains could be triggered to fold and stabilize in response to the correct combination of target-specific recognition and binding events. Obviously, our solvent engineering experiments do not reflect the events or conditions that exist in the enamel matrix during biomineralization, and thus a number of issues cannot be addressed at this time. First, we cannot interpret conformational events that occurred within the intermediate time scale which resulted in loss of structural information within the N-terminus (Fig. 2; Figure S1, Electronic Supporting Information) and these must be addressed by other methods. Second, given the instability of these terminal domains in the monomeric state (Fig. 6), it is possible that, depending upon the nature of the target, secondary structures other than alpha helix may form in these sequences. However, once appropriate targets are identified and experimentally probed, these secondary structure features will be established. Finally, based upon our current understanding of IDP molecular behavior and the persistence of functional, unfolded domains after target binding,[12–14] the fact that the rP172 amelogenin terminal domains can fold does not mean a priori that they will fold for every target that they potentially bind

with. Thus, true binding–folding relationships must be addressed on a case-by-case basis.

Our studies also shed light on the important role that Pro residues play in the folding propensity, conformational exchange, and stability of rP172. On the basis of our TFE experiments, we can categorize rP172 Pro residues into two groups. The first group are the large subset of Pro residues which do not conformationally respond to TFE. These Pro residues primarily reside within the conformationally extended Pro, Gln-containing PPII central domain of rP172 (i.e., I70–P89; P102–P145; Figs. 4–7). Here, it is clear that these Pro groups only experience minor TFE-induced folding transitions. We believe that these Pro residues act as conformational stabilizers and offer improved resistance to TFE-induced folding.[39,40] This resistance to folding may play a role in the function of the PPII containing central domain within rP172, such as the participation in intermolecular contacts that lead to self-assembly at appropriate protein concentrations, or, facilitate the folding and/or stabilization of the adjacent N- and C-termini.[9,39,40,45] The second group of Pro residues are a limited subset (P2, P127, P130, P139, P154, P157, P162), which exhibit significant conformational sensitivity to TFE (Fig. 5). Previous studies have shown that Pro residues can experience *cis-trans* interconversion or function as a molecular "hinge" within polypeptide sequence segments[31,41,42,45] and that TFE affects these parameters.[28,29,33,38] Hence, this responsiveness to TFE indicates that this second group of Pro residues act as folding enhancers and play an important role in the structure and dynamics[31,41,42,45] of the intrinsically disordered protein, rP172. These Pro residues may also be responsible for the increased level of conformational exchange observed in the 70-rP172 sample relative to AQ-rP172 as evidenced by the presence of additional HSQC crosspeaks in 70-rP172 (Figure S1, Electronic Supporting Information).[9] Although we cannot yet identify which sequence regions of rP172 exhibit conformational exchange or what conformational states they are sampling, we do intend to explore this further in future studies.

The "solvent engineering" approach[28] also provides insight into which amino acids participate in the alcohol-induced rP172 oligomerization process (Figs. 2 and 3). We know that TFE weakens protein-protein hydrophobic interactions and promotes inter- and intramolecular hydrogen bonding.[28–33] As shown in Figure 2, the N-terminal P2–W45 sequence region contains a significant number of hydrogen bonding donor-acceptors such as Tyr, Asn, Ser, Tyr, His, and Arg. Interestingly, it is these residues that comprise the "missing" regions of the rP172 HSQC spectra (Fig. 2) that experience intermediate time scale broadening and signal intensity losses. Note that this phenomenon does not occur within the C-terminal domain (Fig. 2). Thus, we propose that the TFE induction of amelogenin oligomerization is initiated by intermolecular sidechain–sidechain hydrogen bonding interactions involving the Tyr, Asn, Ser, Tyr, His, and Arg residues of the N-terminal P2–W45 region. In turn, these interactions lead to relaxational broadening and the loss of corresponding HSQC signals. Obviously, residues within the P2–W45 region which do appear in the HSQC spectra (Fig. 2) correspond to fast time scale motion of amino acids that are not directly involved in TFE-induced intermolecular sidechain-sidechain contacts within the N-terminal region of rP172 oligomers. It will be interesting to learn if intermolecular sidechain hydrogen bonding is also operant during amelogenin oligomerization in aqueous environments as well.

Finally, we wish to place the present work within the context of previous studies of counterion-induced mouse amelogenin dimerization.[15] It was observed that NMR resonances for T21–R31 and Y12–I51 within the N-terminal region underwent intermediate time scale broadening during dimerization, leading to the loss of specific NMR signals, and it was suggested that these sequence regions are responsible for initiating assembly.[15] Interestingly, these counterion-sensitive regions correlate well with the TFE-induced peak broadening effects that were observed for P4-H6, V19-P33, and E40-T58 in rP172 (Fig. 2). Thus, both studies now implicate the N-terminal sequence (the self-assembly A domain) as a site for interprotein contact during *in vitro*-induced self-association. The counterion-induced dimerization studies also reported amide peak broadening for residues within the conserved C-terminal region (L141–T171) at higher counterion concentrations, suggesting that this domain also participates in assembly.[15] However, this peak broadening effect was not observed for the rP172 C-terminal domain in the presence of TFE (Fig. 2). From this, we conclude that counterions can trigger both N- and C-terminal domain protein–protein interactions, but TFE only promotes N-terminal domain interactions. These results are very intriguing, for they show that "solvent" or "ionic" engineering leads to different effects on the intrinsically disordered conformation and assembly of amelogenin. We intend to follow up on this premise in subsequent work.

## Materials and Methods

### *Production of recombinant porcine amelogenin and NMR sample preparation*

Uniformly labeled $^{13}C$, $^{15}N$ recombinant porcine amelogenin [U-$^{13}C$, $^{15}N$ rP172] was expressed in Bio-Express Cell Growth Media (U-$^{13}C$, 98%, U-$^{15}N$, 98%, Cambridge Isotope Laboratories, Lowell, MA) using Escherichia coli strain BL21-codon plus (DE3-

RP, Stratagene, La Jolla, CA) and purified by ammonium sulfate precipitation and reversed phase HPLC as previously described.[9] An unlabeled form of rP172 was also expressed for CD and DLS experiments.

The lyophilized U-$^{13}C$, $^{15}N$ rP172 was used to create two different NMR samples, each 75 micromolar in 300 μL volume. The first, designated as "AQ-rP172," was created by dissolving an appropriate amount of U-$^{13}C$, $^{15}N$ rP172 in Milli-Q pure unbuffered deionized distilled water [90% v/v UDDW/10% v/v $D_2O$ (99.99 atom% D, Cambridge Isotope Laboratories, Lowell, MA)]. This sample is identical to the one utilized in earlier studies of the monomeric low pH form of rP172 and gives rise to a $^{15}N$–$^{1}H$ HSQC spectra that exactly matches the earlier sample.[9] The second, designated as "70-rP172", utilized U-$^{13}C$, $^{15}N$ rP172, UDDW, and an appropriate volume of $D_2$-2,2,2-trifluoroethanol (—OH, —$CD_2$— form, 99.9% atom D, Cambridge Isotope Laboratories, Lowell, MA) to make a 70% v/v TFE/30% v/v UDDW rP172 sample. For NMR experiments we utilized the —$CD_2$- —OH form of TFE which allows hydrogen exchange between the alcohol proton, water protons, and exchangeable NH backbone and sidechain protons on the rP172 molecule. For all NMR samples, after mixing of components the pH was checked with a microelectrode and adjusted with microliter volumes of HCl to a final pH of 3.8. Monitoring of sample pH was performed periodically between experiments and it was found that the sample pH remained stable near 4. Samples were placed in 5 mm symmetrical $D_2O$-matched Shigemi NMR microtubes (Shigemi, Alison Park, PA). No visible aggregation/degradation was evident, as verified by periodic recording of $^{15}N$–$^{1}H$ HSQC experiments.[9]

### *Circular dichroism spectroscopy*

Far UV-CD spectra (260–190 nm) of unlabeled rP172 was recorded on a Jasco J-810 spectropolarimeter after calibrating the instrument with (-) camphorsulfonic acid. The instrument optics were flushed with $N_2$ gas at a flow rate of 20 L/min. For TFE experiments, rP172 was dissolved in cold 50 mM Tris-HCl buffer at pH 5.8 and refrigerated overnight. The solution was then aliquoted to yield 5 micromolar rP172 in the same buffer or appropriate TFE : water ratio (0, 10, 30, 70% v/v) (99.8%, Acros America). For each spectrum, a total of four scans were recorded, averaged, and baseline subtracted. Where appropriate, buffer and TFE subtraction were performed for rP172 samples. All of the TFE experiments were conducted at 25°C using 0.1 cm path length stoppered quartz curvettes.

### *NMR experiments*

Following the protocol utilized in our earlier NMR study of monomeric rP172, we utilized low temperatures (10°C) to slow down protein backbone conformational exchange[9] and minimize aggregation in rP172 samples.[9,34] Protein NMR experiments were conducted on the two labeled rP172 samples using a Bruker AVANCE 800 MHz NMR spectrometer equipped with a triple resonance HCN 5 mm cryoprobe. For the 70-rP172 sample, protein backbone sequential assignments were obtained using an array of multidimensional NMR experiments (HSQC, HNCA, HNCO, HNCACB, HNCACO, HNCOCA).[9–11,35] For the AQ-rP172 sample, $^{15}N$–$^{1}H$-HSQC experiments were performed to qualitatively compare the $^{1}H$ and $^{15}N$ NMR backbone resonance dispersion[34] relative to the 70-rP172 sample and to previously published NMR datasets obtained for monomeric AQ-rP172.[9] NMRPipe software (National Institute of Health, Bethesda, MD) was used to process all NMR data and NMRViewJ (version 8.0.b21 with Java 1.6.0_17, One Moon Scientific, Newark, NJ) was used for sequential assignments. Sparky (SPARKY 3, University of California, San Francisco, CA) was used for spectra visualization. The spectra were referenced with respect to the temperature-corrected water resonance and $^{13}C$ and $^{15}N$ chemical shifts were referenced on the basis of the $^{1}H$ IUPAC guidelines using the unified chemical shift scale.[46] Using NMRViewJ, an input file for calculating the dihedral angle was generated with the respective $^{13}C\alpha$, $^{13}C\beta$, $^{13}CO$, $^{15}N\alpha$ chemical shifts and phi, psi backbone torsion angles for each residue in the 70-rP172 sample using the TALOS program. Using the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts obtained for 70-rP172 and the corresponding published values for AQ-rP172,[9] we calculated the SSP score[43] for these samples. The $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts were internally re-referenced for both data sets.

### *Dynamic light scattering measurements*

DLS was performed at 10 and 22°C using a Wyatt DynaPro Nanostar (Santa Barbara, CA) light scattering instrument. The unlabeled 0% v/v TFE rP172 sample was prepared by diluting 40 μL of unlabeled rP172 stock in UDDW water (2.5 mg/mL) with 10 μL of UDDW water and 50 μL of sodium acetate buffer (50 m*M*, pH 5.5). This resulted in a 1 mg/mL solution of rP172 in 25 m*M* sodium acetate pH 5.5. For the 10% v/v TFE rP172 sample, 10 μL of TFE was introduced to the stock solution instead of 10 μL of water. The results shown in Figure 3 were collected at 22°C, but similar results were obtained at 10°C. The viscosity and refractive index of the TFE:-water mixtures were determined from literature values.[47,48] Efforts to measure 30% and 70% v/v TFE rP172 samples were hindered by the background scatter of those binary solvent mixtures. The data were analyzed using Dynamics 7.0 software and were produced by the program performing a regularization fit using the Dynals algorithm on the

resultant autocorrelation functions. A Rayleigh sphere model was used for the analysis meaning that the hydrodynamic radii calculated were sphere-equivalent radii. By measuring the fluctuations in the laser light intensity scattered by the sample, the instrument is able to detect the speed (diffusion coefficient) at which the particles are moving through the medium. This value is converted to hydrodynamic radius ($R_H$) using the Stokes-Einstein relation:[49]

$$D = \frac{kT}{6\pi\eta R_H}$$

where $D$ is the diffusion coefficient, $k$ is the Boltzmann constant, $T$ is the absolute temperature, $\eta$ is the viscosity, and $R_H$ is the sphere-equivalent hydrodynamic radius.[49]


## Acknowledgments

We thank Prof. Ralf Langen, USC, for use of the CD facilities and Mr. David Maltby, University of California, San Francisco, for the mass spectrometry analysis of the labeled amelogenin under Project # 403. We also wish to thank Dr. Michael Goger of NYSBC for his help with NMR experiments. This article represents contribution number 58 from the Laboratory for Chemical Physics, New York University.


## References


1. Lowenstam HA, Weiner S (1989) On biomineralization. New York: Oxford University Press.
2. Nanci A, Enamel: composition, formation, and structure. In: Nanci A, Ed. (2008) Ten cate's oral histology development, structure, and function. St. Louis: Mosbey Elsevier, pp 141–190.
3. Moradian-Oldak J, Paine ML, Mammalian enamel formation. In: Sigel A, Sigel H, Sigel RKO, Eds. (2008) Metal ions in life sciences. Chichester: Wiley, pp 507–546.
4. Fincham AG, Moradian-Oldak J, Simmer JP (1999) The structural biology of the developing dental enamel matrix. J Struct Biol 126:270–299.
5. Wiedemann-Bidlack FB, Beniash E, Yamakoshi Y, Simmer JP, Margolis HC (2007) pH triggered self-assembly of native and recombinant amelogenins under physiological pH and temperature in vitro. J Struct Biol 160:57–69.
6. Ijima M, Moradian-Oldak J (2004) Control of octacalcium phosphate and apatite crystal growth by amelogenin matrices. J Mater Chem 14:2189–2199.
7. Beniash E, Simmer JP, Margolis HC (2005) The effect of recombinant mouse amelogenins on the formation and organization of hydroxyapatite crystals in vitro. J Struct Biol 149:182–190.
8. Moradian-Oldak J, Lakshminarayanan R (2010) Intrinsic disorder in amelogenin. In: Amelogenins: multifaceted proteins for dental and bone formation and repair, Michel Goldberg, Ed. Bentham Science Publishing, Ltd., London, UK, Chapter 10. pp 106–132.
9. Delak K, Harcup C, Lakshminarayanan R, Zhi S, Fan Y, Moradian-Oldak J, Evans JS (2009) The tooth enamel protein, porcine amelogenin, is an intrinsically disordered protein with an extended molecular configuration in the monomeric form. Biochemistry 48: 2272–2281.
10. Dyson HJ, Wright PE (1998) Equilibrium NMR studies of unfolded and partially folded proteins. Nat Struct Biol 7:499–503.
11. Dyson HJ, Wright PE (1998) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. Meth Enzymol 339:258–270.
12. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41:415–427.
13. Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci 11:739–756.
14. Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27:527–533.
15. Buchko GW, Tarasevich BJ, Bekhazi J, Snead ML, Shaw WJ (2008) A solution NMR investigation into the early events of amelogenin nanosphere self-assembly. Biochemistry 47:13215–13222.
16. Jin T, Ito Y, Luan X, Dangaria S, Walker C, Allen M, Kulkarni A, Gibson C, Braatz R, Liao X, Diekwisch TGH (2009) Elongated polyproline motifs facilitate enamel evolution through matrix subunit compaction. PLOS Biol 7:1–10.
17. Lakshminarayanan RY, Hegde BG, Du C, Fan D, Moradian-Oldak J (2009) Analysis of secondary structure and self-assembly by variable temperature circular dichroism and isothermal titration calorimetry. Protein Struct Funct Bioinform 76:560–569.
18. Lakshminarayanan R, Fan D, Du C, Moradian-Oldak J (2007) The role of secondary structure in the entropically driven amelogenin self-assembly. Biophys J 93: 3664–3674.
19. Matsushima N, Izumi Y, Aoba T (1998) Small angle x-ray scattering and computer aided molecular modeling studies of 20 kDA fragment of porcine amelogenin. Does amelogenin adopt an elongated bundle structure? J Biochem 123:150–156.
20. Paine ML, Lei YP, Dickerson K, Snead ML (2002) Altered amelogenin self-assembly based on mutations observed in human X-linked amelogenesis imperfecta (AIH1). J Biol Chem 277:17112–17116.
21. Zhu DH, Paine ML, Luo W, Bringas P, Snead ML (2006) Altering biomineralization by protein design. J Biol Chem 281:21173–21182.
22. Bartlett JD, Ganss B, Goldberg M, Moradian-Oldak J, Paine ML, Snead ML, Wen X, White SN, Zhou YL (2006) Protein-protein interactions of the developing enamel matrix. Curr Top Dev Biol 74:57–115.
23. Yang X, Wang L, Qin Y, Sun Z, Henneman ZJ, Moradian-Oldak J, Nancollas GH (2010) How amelogenin orchestrates the organization of hierarchical elongated microstructures of apatite. J Phys Chem B 114:2293–2300.
24. Gibson CW (2008) The amelogenin "enamel proteins" and cells in the periodontium. Crit Rev Eukaryot Gene Expr 18:345–360.
25. Veis A (2003) Amelogenin gene splice products: potential signaling molecules. Cell Mol Life Sci 60:38–55.
26. Fincham AG, Moradian-Oldak J (1993) Amelogenin post-translational modification: carboxy-terminal processing and the phosphorylation of bovine and porcine "TRAP" and "LRAP" amelogenins. Biochem Biophys Res Commun 197:248–255.
27. Laksminarayanan R, Bromley KM, Lei YP, Snead ML, Moradian-Oldak J (in press) Perturbed amelogenin secondary structure leads to uncontrolled aggregation in

amelogenesis imperfecta mutant proteins. 285:40593–40603.
28. Buck M (1998) Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. Quart Rev Biophys 31:297–355.
29. Maroun RG, Krebs D, El Antri S, Deroussent E, Lescot E, Troalen F, Porumb H, Goldberg ME, Fermandijan S (1999) Self-association and domains of interaction of an amphipathic helix peptide inhibitor of HIV-1 integrase assessed by analytical ultracentrifugation and NMR experiments in trifluoroethanol/water mixtures. J Biol Chem 274:34174–34185.
30. Cammers-Goodwin A, Allen TJ, Oslick SL, McClure KF, Lee JH, Kemp DS (1996) Mechanism of stabilization of helical conformations of polypeptides by water containing trifluoroethanol. J Am Chem Soc 118: 3082–3090.
31. Ragona LM, Catalano M, Zetta L, Longhi R, Fogolari F, Molinari H (2002) Peptide models of folding initiation sites of bovine beta-lactoglobulin: identification of nativelike hydrophobic interactions involving G and H strands. Biochemistry 41:2786–2796.
32. Luo P, Baldwin RL (1997) Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. Biochemistry 36: 8413–8421.
33. Hong DP, Hoshino M, Kuboi R, Goto Y (1999) Clustering of fluorine-substituted alcohols as a factor responsible for their marked effects on proteins and peptides. J Am Chem Soc 121:8427–8433.
34. Li M, Liu J, Ran X, Fang M, Shi J, Qin H, Goh JM, Song J (2006) Resurrecting abandoned proteins with pure water: CD and NMR studies of protein fragments solubilized in salt-free water. Biophys J 91:4201–4209.
35. Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu G, Ramelot TA, Eletsky A, Szyperski T, Kennedy MA, Prestegard J, Montelione GT, Baker D (2010) NMR Structure determination for larger proteins using backbone-only data. Science 327: 1014–1018.
36. Wishart DS, Case DA (2001) Use of chemical shifts in macromolecular structure determination. Meth Enzymol 338:3–34.
37. Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 222: 311–333.
38. Tamburro AM, Bochicchio B, Pepe A (2003) Dissection of human tropoelastin: exon-by-exon chemical synthesis and related conformational studies. Biochemistry 42: 13347–13362.
39. Darnell G, Orgel JPRO, Pahl R, Meredith SC (2007) Flanking polyproline sequences inhibit beta-sheet structure in polyglutamine segments by inducing PPII-like helix structure. J Mol Biol 374:688–704.
40. Cubelis MV, Caillez F, Blundell TL, Lovell SC (2005) Properties of Polyproline II, a secondary structure element implicated in protein-protein interactions. Protein Struct Funct Bioinform 58:880–892.
41. Bhattacharyya R, Chakrabarti P (2003) Stereospecific interactions of proline residues in protein structures and complexes. J Mol Biol 331:925–940.
42. Dumy P, Keller M, Ryan DE, Rohwedder B, Wohr T, Mutter M (1997) Pseudo-prolines as a molecular hinge: reversible induction of cis amide bonds into peptide backbones. J Am Chem Soc 119:918–925.
43. Marsh JA, Singh VK, Jia J, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha and gamme-synuclein: implications for fibrillation. Protein Sci 15: 2795–2804.
44. Fan D, Du C, Sun Z, Lakshminarayanan R, Moradian-Oldak J (2009) In vitro study on the interaction between the 32 kDa enamelin and amelogenin. J Struct Biol 166:88–94.
45. Butcher DJ, Nedved ML, Neiss TG, Moe GM (1996) Proline pipe helix: structure of the Tus proline repeat determined by 1H NMR. Biochemistry 35:698–703.
46. Wang Y, Wishart DS (2005) A simple method to adjust inconsistently referenced 13C and 15N chemical shift assignments of proteins. J Biomol NMR 31:143–148.
47. Lin CT, Marques ADS, Pessine FBT, Guimarães WON (1981) The shape of the refractive index versus composition curves for hydrogen-bonded liquid mixtures. J Mol Struct 73:159–169.
48. Palepu R, Clarke J (1989) Viscosities and densities of 2,2,2-trifluoroethanol + water at various temperatures. Thermochim Acta 156:359–363.
49. Schärtl W (2007) Light scattering from polymer solutions and nanoparticle dispersions, 1st edition. Germany: Springer-Verlag, Heidelberg.